REMARKS

This Amendment is being filed in response to the Office Action mailed November 20, 2003. Reconsideration and allowance of all of the claims is earnestly solicited.

Claims 38-57 and 59-62 remain in the application. No claim has been allowed.

Claims 41, 42 and 61 were objected to as being dependent upon a rejected base claim. The Examiner indicated, however, that these claims would be allowable if rewritten in independent form, including all the limitations of the base claim and any intervening claims.

Claim 41 is now rewritten as an independent claim and should be allowable.

Claim 42 depends directly from claim 41 and is now also allowable for the same reason.

Claims 43 through 46 either depend directly or indirectly from Claim 41 and are likewise patentable.

Claim 61 is now also rewritten in independent form and is likewise allowable.

The remaining claims stand rejected under 35 U.S.C. 103 (a) as being unpatentable over Ahmed (U.S. Patent No. 6,263,50), in view of Vora, et al. (U.S. Patent No. 5,819,273) and Ahn (U.S. Patent No. 5,696,963).

The Examiner argues that Ahmed discloses a system for categorizing a news story into subject matter categories (such as sports, news, travel, etc.), and which also allows a user to search the text of news stories. This characterization of Ahmed is correct. However, we cannot agree that the claimed "extraction" of "named entities" is the equivalent of just any "keyword" "search". Furthermore, the Applicant's claim requires not just a ranked presentation of search results from among a group of documents, but rather requires presenting a list of extracted named entities and their corresponding frequency of occurrence across many story segments. For these reasons, the Examiner has failed to make a prima facie case of obviousness.

More particularly now, the Examiner argued at page 3, lines 8-10, of the last response that

"...It should be noted that the sports, travel, computers, international news, etc., meet the claimed "named entities" i.e., location, sports or travel is a "named entity".

However, we cannot agree. Users of computerized search systems often have information needs that are not only about a specific category of information, that is "sports", "travel", "computers", etc., but rather about specific people, organizations and locations that occur in stories over a time period. For example, a user who wishes to retrieve stories about "sports" would merely retrieve a list of many different stories about different types of sports and different sports figures.

This is quite different from a user interested in querying for stories only about certain "named entities" such as the sports figure "Wayne Gretzky" or "Tiger Woods". The latter case also exhibits the value of named entities and their difference from search terms in general. If, for example, a user performs a typical web type search engine query using the term "Tiger Woods", the user retrieves stories not only about the famous golf athlete (the person), but also stories about "tigers", the wild animals, and also many articles about places having lots of trees.

When a search query specifies a logical operator and searches for the terms "Tiger" AND "Woods", the user will retrieve articles about the person "Tiger Woods", but will also retrieve articles about tiger animals who might be found roaming in the woods. Only when the search task is sophisticated enough to recognize the phrase "Tiger Woods" as a named entity (person), does it retrieve only articles about the famous golfer, and no others.

Categories of information, such as "sports" or "travel" are also not "named entities" in the manner in which they are used in the applicant's specification. Rather these are simply broad categories of information. There are lots of different "kinds" of sports, that is football, basketball, baseball, hockey, soccer and the like and lots of individual players and locations such as specific arenas associated with each of those.

We also ask the Examiner to more carefully consider what the claim actually states. Clause (b) of Claim 59 is a step of named entity extraction, that is, a method for extracting a specific information structure, including proper names, time, and for numerical expressions from written or spoken languages. Named entity extractions has been recognized as a specific branch of linguistics and intensively investigated in the past several years. For example, the National Institute for Science and Technology (NIST) has for several years held a annual competition to evaluate algorithms to extract named entities from text, called the Text Evaluation and Retrieval Competition and Message Understanding Conference (MLC). A corresponding definition of

"named entities" and their use as event queues in text streams, including named entities such as "persons, locations and organizations" was defined in the Applicants' specification. It was described, for example, at page 11 of the Applicants' specification that the extraction of named entities from

> "a text stream 108, such as <u>people, places and things</u> is accomplished by detecting clue words and other linguistic indicators of named entities. It can include capitalization clues but are not limited to capitalization clues."

A reference to a specific preferred tool for automatic named entity extraction was given at page 15 of the specification, i.e., the cited paper by Aberden, J., entitled "Description of the Alembic System Used for MUC-6".

Thus, <u>automatic named entity</u> extraction from text <u>is not the equivalent of a category definition</u>.

Furthermore, named entity extraction from is not the equivalent of a simple search for a known keyword. Searching of documents for keywords requires a search term to be provided by a user. For example, in a typical web browser search, the user provides the search term himself.

In contrast to this, step (b) of Claim 59 requires <u>extracting named entities from a text information stream</u>. It is making reference to a automated process for extracting the names of persons, places and organizations, whatever they may be, from a text stream. This is of course, quite different from the cited art since the "named entities" are not known in advance–they are found by the extraction process.

Claim 59 is thus not merely directed to "searching" for a particular word known in advance. Rather, its clause (b) calls for determining all words of a particular class (named entities) that exist in the story text.

Furthermore, Claim 59 goes on to require:

"©) using the extracted named entities as search criteria to select from among a plurality of story segments"

This feature is also not taught in Ahmed. Ahmed does allow a user to search news stories using keywords, and Ahmed also discloses methods for categorizing news stories into subject matter categories.

However, this is not the same thing as what the claim states. Clause ©) of claim 59 requires using the <u>extracted named entities</u> as a search criteria to select a story segment. In other words, after analyzing a news story for named entities, the Applicants' invention then allows the user to view a list of different extracted entities, and then further select story segments related to the extracted entities.

The claim even goes on to further require a step of:

"(d) presenting a list of named entities and their corresponding frequency of occurrence in story segments over a selected time period"

Preferred embodiments of this feature of Applicants' invention are shown in Figures 14 and 18 of the Application. In the case of Figure 14, the presenting steps show a graph of the named entities extracted from story segments and their frequency of occurrence. The named entities extracted from stories in this example here are "Mother Theresa", "White House", "Princess Diana", "Washington", "Clinton", and "Army". These named entities <u>were not provided by the user in advance, but rather these were automatically determined from story segments</u> in the database.

A frequency of occurrence plot is associated with each of a number of extracted named entities. The frequency of occurrence is used as criteria for the user to select story segments for a more in-depth look at the news. As expressed in dependent claims 38 and 39, and in the description of Figure 14, it is evident that the user can use the frequency of named entities graph. The user selects a named entity in which further interest is shown, by "clicking on" the named entity "Princess Diana", and the system then retrieves a list of stories about Princess Diana in a selected time period.

Figure 18 provides a similar view of extracted named entities, formatted as a table rather than a graph. Again, there is a presentation of extracted named entities and a means (that is the hyperlinks shown in the value column), to select story segments associated with a selected one of the extracted named entities.

It is clear to us that Ahmed does not teach anything at all about extraction of named entities and certainly nothing at all about their use in selection of further story segments for presentation. Claim 59 and certainly dependent claims 38 and 39 should be allowed.

The Examiner, in fact, admits that Ahmed fails to disclose presenting a number of occurrences in story segments, and then goes on to cite Vora and Ahn for that proposition. However, this portion of the Examiner's argument also contains at least two logical inconsistencies. First of all, the claim language does not merely require "presenting the number of occurrences of a search term in a story segment". Rather the claim language recites using the frequency of extracted named entities <u>as criteria to select from among a plurality of story segments</u>. The two are not the same thing.

More particularly, the Examiner cites column 10, lines 49-65 of Vora for the proposition that the prior art ranks documents according to the number of keywords present. As described in Vora, a list of documents is returned with a number indicating a frequency of occurrence of a particular keyword in each document. However, the Applicants' claim requires, in clause (d), a step of <u>presenting a list of named entities</u> and then a number indicative of the frequency of occurrence of the named entities. The number indicated is not just for a single document but plural <u>documents</u>. Thus, Applicants' claim requires determining the total number of occurrences of <u>named entities</u> throughout a database of story segments. Thus the graphs shown in Figures 14 and 18 of Applicants' specification are totaling the total number of times that the named entity has been listed, not in just one document, but a number of occurrences in many story <u>segments.</u>

Stated another way, the Applicants' claim is not directed to listing how many times a given word appears in a document (leaving aside for a moment the distinction between named entities and search terms presented by a user), but rather is presenting a list showing the frequency of occurrence of the named entity word across multiple story segments in a database. Using the example above, a user of the Applicants' system sees "Princess Diana" as an extracted named entity, with a total number of times across all stories, that "Princess Diana" appears. This display is presented along with other named entities in the same time period, such as "White House" or "Pentagon". With a system that is a combination of Ahmed and Vora, as urged by the Examiner, the user would merely see a list of documents that mention "Princess Diana", with the document that mentions "Princess Diana" most often listed at the top of the search results.

So with the combination of Ahmed and Vora, one would be left with a system that merely counts the number of times that a search term specified by a user appears in each of many news stories, and then ranks the news stories by frequency of occurrence of the search term. In

Applicants' claimed invention, news stories are first searched for <u>any</u> named entities that might occur within. These named entities are then ranked across multiple stories, not each story itself.

The Examiner has thus not set forth a prima facie case of obviousness, since all elements of the claim are also not found in Ahmed or Vora.

The Examiner also cites Ahn, stating that Ahn "teaches displaying the number of hits of keywords appearing in a document (sic) search by user." We do admit that the Abstract of Ahn teaches storing a number of hits in a group hits table with each hit entry corresponding to a different document in which a predetermined keyword appears. The system of Ahn thus determines the number of times a keyword appears in each document. Beyond the Abstract, it is evident from reading Ahn (and specifically looking at Figures 2, 4 and 5 of Ahn), that he is developing a hit table as a list of keywords, a document ID, and a number of hits in each document.

However, Ahn is not showing or ranking the <u>total</u> number of "hits" of one search term against other search terms; he is certainly not allowing or teaching one of skill in the art to <u>extract named entities</u> and/or display a summary of a total number of times that named entities appear in a group of stories. Rather, with Ahn, the user must start by giving the system a keyword to search for. In Applicants' invention, the system automatically extracts named entities from multiple story segments and then determines a total number of extracted named entities across all stories, to then permit a user to select a subject of interest.

Putting it another way, with the Applicants' invention, the user need not decide in advance that he or she is interested in "Princess Diana" stories. Rather, the claimed system permits automated determination of current "hot topics" in multiple news broadcasts, by extracting named entities. These are then ranked in a summary form, without the user having to know in advance what those key search terms are. The user can select the most commonly mentioned words.

We also respectfully submit that one of skill in the art would not have looked to Ahn or Vora to augment the teachings of Ahmed. That is, Ahn and Vora are merely attempting to rank relevant documents by counting up a number of times that a particular word appears in a document. Applicants' invention, on the other hand, is looking for a way to summarize news

broadcasts by using automated named entity extraction tools, and then ranking the results by the number of times a named entity occurs among a large number of stories.

Claims 59 should therefore be allowed.

With regard to Claim 38, since no cited prior art attempts to extract named entities or even discusses the extraction of named entities whatsoever, the prior art also does not teach extracting text information from story segments, and then linking that in a display together with named entity data for that story segment.

For example, Applicants' system, as shown in Figs. 17 and 19, is capable of showing a browser type display that not only shares information extracted from a news story (such as text and a key frame), but also named entities extracted from that story. Thus, in Fig. 17, the invention has automatically extracted the named entity "Pentagon"; and with the example of Figure 19, the named entity "Zaire" has been extracted and presented as a summary display together with a key frame and other information extracted from the text of the story.

If one were to combine the teachings of Ahmed, Ahn and/or Vora, one would simply arrive at a system wherein one would have text data, summary data, and perhaps key terms that were previously specified by a user shown together. However, named entity data as extracted from the stories, would not be presented.

The Examiner has also continued in the rejection of Claim 40. However, Claim 40, which depends from Claim 59, requires extracting story summary data using the named entities as a basis. No discussion is found in the cited prior art of named entities in the first place, and thus no suggestion of extracting summary data using a named entity as a basis can be found.

We do also note that the Examiner has allowed Claim 41, which adds to Claim 40 a further definition of what is extracted from the story segment (that is a frequency of occurrence). Certainly, Claim 40 should be allowable for the same reasons that the Examiner has allowed Claim 41.

Claims 55 through 57 were also rejected under Ahmed, Ahn and/or Vora under 35 U.S.C. 103(b). Claim 55 would be allowable for the same reasons given above for Claim 59.

Furthermore, at least Claim 56 should also be allowed. In particular, the prior art does not teach generation of a summary representation including extracted named entities.

CONCLUSION

In view of the above amendments and remarks, it is believed that all claims are now in condition for allowance, and it is respectfully requested that the application be passed to issue. If the Examiner feels that a telephone conference would expedite prosecution of this case, the Examiner is invited to call the undersigned.

Respectfully submitted,

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

By _____

David J. Thibodeau, Jr.
Registration No. 31,671
Telephone: (978) 341-0036
Facsimile: (978) 341-0136

Concord, MA 01742-9133
Dated: 5/20/04